

Theoretical and Empirical Standard Errors for Two Population Invariance Measures in the Linear Equating Case

*Alina A. von Davier
Jonathan R. Manalo
Frank Rijmen*

April 2008

ETS RR-08-24



**Theoretical and Empirical Standard Errors for Two Population Invariance Measures in
the Linear Equating Case**

Alina A. von Davier, Jonathan R. Manalo, and Frank Rijmen
ETS, Princeton, NJ

April 2008

As part of its educational and social mission and in fulfilling the organization's nonprofit charter and bylaws, ETS has and continues to learn from and also to lead research that furthers educational and measurement research to advance quality and equity in education and assessment for all users of the organization's products and services.

ETS Research Reports provide preliminary and limited dissemination of ETS research prior to publication. To obtain a PDF or a print copy of a report, please visit:

<http://www.ets.org/research/contact.html>

Copyright © 2008 by Educational Testing Service. All rights reserved.

ETS, the ETS logo, and LISTENING. LEARNING.
LEADING. are registered trademarks of Educational Testing
Service (ETS).



Abstract

The standard errors of the 2 most widely used population-invariance measures of equating functions, root mean square difference (RMSD) and root expected mean square difference (REMSD), are not derived for common equating methods such as linear equating. Consequently, it is unknown how much noise is contained in these estimates. This paper describes 2 methods for obtaining the standard errors for RMSD and REMSD. The delta method relies on an analytical approximation and provides asymptotic standard errors. The grouped jackknife method is a sampling-based method. Both methods were applied to a real data application. The results showed that there was very little difference between the standard errors found by the 2 methods.

Key words: Population invariance, RMSD, REMSD, jackknife method, standard error

Acknowledgments

The authors would like to thank Shelby Haberman and Paul Holland for their thoughtful suggestions and comments that helped lay the foundation for this project.

Introduction

Test equating methods are used to produce scores that are interchangeable across different test forms. Lord (1980) and others (see Kolen & Brennan, 2004, for an overview) state several requirements for equating. One of these requirements is the *population invariance requirement*, which states, “The choice of (sub) population scores of tests X and Y should not matter. In other words, the equating function used to link the scores of X and Y should be population invariant” (Dorans & Holland, 2000, p. 28). As discussed by Dorans (2003), population invariance is not only important in and of itself, but it also plays a key role in another equating requirement, the *equity requirement*, which states, “It ought to be a matter of indifference for an examinee to be tested by either one of two tests that have been equated” (Dorans & Holland, p. 28). In short, population invariance and equity are interrelated, and equity in equating cannot be achieved if population invariance does not hold. That is, the tests can only meet the equity requirement and be equatable if it is not only irrelevant which test (of those being equated) the examinee takes, but also which (sub)population of scores is used to obtain the equating function.

Dorans and Holland (2000) proposed using two measures to determine if the population-invariance requirement is met. One measure is the root mean square difference (RMSD), which is defined as the RMSD between the subpopulation linking functions and the overall linking function. That is, if scores y on test Y (i.e., the new form) are equated to scores x on test X (i.e., the old form) on a target population of examinees P , then

$$\text{RMSD}(y) = \frac{\sqrt{\sum_c w_c [e_{P_c}(y) - e_P(y)]^2}}{\sigma_{XP}}, \quad (1)$$

where e_{P_c} is the equating function for subpopulation P_c ($c = 0, 1, 2, \dots, C - 1$) in population P , e_P is the equating function for the overall population P , w_c is the weight assigned to subpopulation P_c , and σ_{XP} is the standard deviation of X for population P . (With some abuse of notation, X and Y denote both the test and the random variable representing the score on the test in this example.)

The RMSD is interpreted as an effect size is (Dorans & Holland, 2000). Specifically, it is the amount of invariance of the linking function for subpopulation c to the overall linking function. So, a value of .15 for $\text{RMSD}(y)$, say $y = 80$, is interpreted as an RMSD of 15% of the standard deviation of test X in P in the linking functions at score 80 of test Y .

Another measure of population invariance is the root expected mean square difference (REMSD), which is defined as a summative measure of the values of the $\text{RMSD}(y)$

$$\text{REMSD} = \frac{\sqrt{\sum_c w_c E_P \left\{ \left[e_{P_c}(Y) - e_P(Y) \right]^2 \right\}}}{\sigma_{XP}}, \quad (2)$$

where $E_P\{\cdot\}$ is the expected value over Y .

The REMSD is interpreted similarly to the RMSD. For example, if $\text{REMSD} = .30$, then the RMSD in the linking functions across the scores of test Y is, on average, 30% of the standard deviation of test X in P .

The extent to which equating functions are population-invariant is commonly addressed in terms of differences that matter (DTM), the differences in equated scores that are greater than would be corrected for by the score rounding that occurs before score reporting (Dorans & Feigenbaum, 1994). The DTM criterion is an indication of practical, rather than statistical, significance of equating function differences (see von Davier and Wilson, 2006, for a discussion of the various criteria for investigating population-invariance indices). If the population-invariance requirement is met, then both the RMSD and REMSD should be lower than the standardized DTM (SDTM), which is the DTM divided by the standard deviation of the old form score (von Davier & Wilson; Dorans & Feigenbaum). This approach is illustrated in recent papers by von Davier, Holland, and Thayer (2004b); Dorans, Holland, Thayer, and Tateneni (2003); and Yang, Dorans, and Tateneni (2003).

When evaluating the RMSD and REMSD, their magnitude is not all that matters. Information regarding the accuracy of the population invariance estimates is a valuable source of information as well. It tells us the degree to which we can trust the estimates of the RMSD and REMSD when deciding whether or not an equating function is fair and population-invariant. That is, are the estimates close to their population values, or is there substantial noise in the

estimates? One way to answer this question is to examine the standard errors (i.e., the noise) of the RMSD and REMSD. Little research has been conducted in this respect. A notable exception is Moses (2006), which applied the delta method to obtain the asymptotic standard errors for the kernel equating (KE) method in an equivalent-group (EG) equating design.

A first goal of this paper is to derive the asymptotic standard errors of the RMSD and REMSD for the linear equating method in the EG design. Even though the linear equating method can be seen as a limiting case of the KE method, and so in principle the results of Moses (2006) could be used, we think it is worthwhile to present the derivations separately for the widely used linear equating function. As in Moses (2006), the standard errors are derived by the delta method.

A second goal of this paper is to present the sampling-based grouped jackknife method (Efron, 1982; Efron & Tibshiriani, 1993) as an alternative way of estimating the standard errors of the RMSD and REMSD. A main advantage of sampling-based methods is that no complex derivations are involved. Even for the relatively simple case of linear equating in the EG design, the analytical derivations become quite tedious using the delta method. If the results of the jackknife method were close to the results of the delta method, the former could be considered to be a viable alternative to the delta method for computing the standard errors of the RMSD and REMSD in more complex designs.

The remainder of the paper is organized as follows. After a short review of the linear equating method, the delta method and the grouped jackknife technique for obtaining the standard errors of the RMSD and REMSD are described. For ease of exposition, the methods are described for the case of two subpopulations, but the results can be generalized to more than two subpopulations in a straightforward way. Subsequently, both methods are illustrated with a real data application. The effect of sample size is investigated in the following section. Conclusions are drawn in the last section.

The Linear Equating Method

In the linear equating method, Y is equated to X scores using a linear equating function. For each subpopulation c , the equating function is

$$e_{P_c}(y) = \mu_{XP_c} + \frac{\sigma_{XP_c}}{\sigma_{YP_c}}(y - \mu_{YP_c}), \quad (3)$$

where μ_{XP_c} and σ_{XP_c} are the mean and standard deviation of X in subpopulation c , and μ_{YP_c} and σ_{YP_c} are the mean and standard deviation of Y in the subpopulation.

The population linear equating function is obtained by replacing the subpopulation means and standard deviations with their population counterparts (denoted without the subpopulation subscript c):

$$e_P(y) = \mu_{XP} + \frac{\sigma_{XP}}{\sigma_{YP}}(y - \mu_{YP}), \quad (4)$$

In practice, the (sub)population means and standard deviations are unknown, and linear equating proceeds by evaluating (3) and (4) in the estimated means and variances.

Asymptotic Standard Errors

Asymptotic standard errors can be derived for the RMSD and the REMSD of the linear equating function in the EG design using the delta method (Bishop, Feinberg & Holland, 1975; Kendall & Stuart, 1977). The delta method is an analytical method that can be summarized by saying that, if a vector of parameter estimates $\hat{\pi}$ is asymptotically normally distributed (such as maximum likelihood estimates under fairly general conditions), then a smooth function f of these parameter estimates is also asymptotically normally distributed, with asymptotic variance-covariance matrix

$$\text{Var}(f(\hat{\pi})) \cong \mathbf{J}_f(\pi) \Sigma(\pi) \mathbf{J}_f'(\pi), \quad (5)$$

where $\mathbf{J}_f(\pi)$ is the Jacobian (i.e., the matrix of the first derivatives of the smooth function f with respect to the components in π). In practice, the Jacobian is evaluated at the maximum likelihood estimates of π .

For the linear equating function in the EG design and in the case of two subpopulations P_0 and P_1 , the components of the vector parameter π are the means and variances of the two variables X and Y in each of the two subpopulations, that is

$$\pi = \left(\mu_{X_0}, \sigma_{X_0}^2, \mu_{Y_0}, \sigma_{Y_0}^2, \mu_{X_1}, \sigma_{X_1}^2, \mu_{Y_1}, \sigma_{Y_1}^2 \right)', \quad (6)$$

and the role of the function f in (5) above is played by the RMSD and REMSD indices for the linear equating function. The entries for $\mathbf{J}_f(\boldsymbol{\pi})$ are given in the appendix.

$\boldsymbol{\Sigma}$ is the 8×8 variance-covariance matrix of the vector parameter $\boldsymbol{\pi}$ and can be written as

$$\boldsymbol{\Sigma} = \begin{pmatrix} \begin{pmatrix} \text{Var}(\pi_{X0}) & \mathbf{0} \\ \mathbf{0} & \text{Var}(\pi_{Y0}) \end{pmatrix} & \mathbf{0} \\ \mathbf{0} & \begin{pmatrix} \text{Var}(\pi_{X1}) & \mathbf{0} \\ \mathbf{0} & \text{Var}(\pi_{Y1}) \end{pmatrix} \end{pmatrix}. \quad (7)$$

The off diagonal contains $\mathbf{0}$ matrices because we assumed that the subpopulations are partitions of P and that the groups were randomly drawn from P . $\text{Var}(\pi_{Xc})$ and $\text{Var}(\pi_{Yc})$ are the covariance matrices of the components of the parameter $\boldsymbol{\pi}$ for subpopulation P_c for tests X and Y , respectively. More precisely, these matrices are represented as

$$\text{Var}(\pi_{Xc}) = \begin{pmatrix} \text{Var}(\mu_{Xc}) & \text{Cov}(\mu_{Xc}, \sigma_{Xc}^2) \\ \text{Cov}(\sigma_{Xc}^2, \mu_{Xc}) & \text{Var}(\sigma_{Xc}^2) \end{pmatrix}, \quad (8)$$

$$\text{Var}(\pi_{Yc}) = \begin{pmatrix} \text{Var}(\mu_{Yc}) & \text{Cov}(\mu_{Yc}, \sigma_{Yc}^2) \\ \text{Cov}(\sigma_{Yc}^2, \mu_{Yc}) & \text{Var}(\sigma_{Yc}^2) \end{pmatrix}. \quad (9)$$

The expressions for the entries for the matrices in (8) and (9) can be found in Kolen (1985; Table 2, estimates for the non-normal case).

The standard errors of the RMSD and REMSD are the square root of the variance. However, conditions for applying the delta method may not be fulfilled everywhere because the RMSD (REMSD) is not differentiable at 0, its value under the null hypothesis. This can be seen in the entries of $\mathbf{J}_f(\boldsymbol{\pi})$, where the RMSD (REMSD) appears in the denominator. Note that the delta method requires that the equating functions be continuous and differentiable almost everywhere (Rao, 1973). The RMSD and REMSD functions are differentiable except on some

finite points with total probability of 0. This means that this requirement of the delta method is actually met in this situation.

Sampling-Based Standard Errors

Standard errors for RMSD and REMSD can also be obtained using the grouped jackknife method. Because the method is sampling-based, it can be easily generalized beyond the EG equating design and the linear equating functions.

Suppose that one is interested in the standard error of a statistic T computed on a sample of size N , but the reference distribution of T is unknown. Alternatively, the reference distribution of T may be known asymptotically, but N may not be large enough to rely on an asymptotic result. In our case, the RSMD and REMSD play the role of T .

The grouped jackknife technique is, like the bootstrap, a sampling-based technique that can be used to obtain standard errors for a statistic T whose reference distribution is unknown. In sampling-based techniques, multiple samples are drawn from the original dataset, T is estimated in each of these samples, and the variation of these estimates over the samples is used to approximate the true standard error of the T (Efron, 1982; Efron & Tibshirani, 1993).

In the bootstrap technique, samples of size N are drawn with replacement from the original sample, whereas in the grouped jackknife method, g groups of size h , with $N = gh$, are drawn without replacement from the original dataset. That is, the sample is randomly split into g groups of size h . Grouped jackknife standard errors are then obtained as follows:

1. Compute the statistic of interest, T_i for the i^{th} group ($i = 1, \dots, g$), which is the total group with the i^{th} block of h observations removed; and
2. Calculate the variance of the estimator as

$$\text{Var}(T) = \frac{g-1}{g} \sum_{i=1}^g (T_i - \bar{T})^2, \quad (10)$$

$$\text{where } \bar{T} = \frac{1}{g} \sum T_i. \quad (11)$$

In this report, the statistics of interest are the RMSD and REMSD.

The larger the number of groups g that is chosen, the more precise the estimate of the standard error will be, but computation time will increase as well.

As was the case for the delta method, conditions for applying the jackknife method may not be fulfilled everywhere due to the fact that the RMSD is not differentiable at 0, its value under the null hypothesis.

In the next section, both the delta and grouped jackknife methods for obtaining the standard errors of the RMSD and REMSD are applied to a real data set.

Real Data Application: A Professional Licensure Study

The data consisted of two randomly equivalent samples (i.e., an EG design) that took one of two forms (*X* and *Y*) of a professional licensure examination (consisting of 40 dichotomously scored items). In the operational administration, Forms *X* and *Y* were spiraled to achieve randomly equivalent samples. For the purposes here, the subpopulations of interest were chosen to be males and females. Examinees who did not provide their gender were removed from the data. Table 1 provides the descriptive statistics of the final data. As shown, slightly more examinees took *Y* than *X*, and roughly four times more females than males took each form.

In general, Forms *X* and *Y* have similar negatively skewed distributions (Figures 1 and 2), but Form *Y* was slightly easier than Form *X*. For both forms, males performed slightly better than females.

Table 1

Descriptive Statistics for the Raw Scores of Forms X and Y

Form	Group	<i>N</i>	Mean	Std. dev.	Minimum	Maximum	Skew	Kurtosis
<i>X</i>	Female	4,221	29.16	7.17	6.00	40.00	-0.64	-0.29
	Male	1,186	30.25	7.12	6.00	40.00	-0.86	0.11
	Total	5,407	29.40	7.17	6.00	40.00	-0.68	-0.23
Reliability		.88						
<i>Y</i>	Female	4,250	30.47	7.27	4.00	40.00	-0.89	0.03
	Male	1,139	31.50	7.07	8.00	40.00	-1.01	0.34
	Total	5,389	30.69	7.24	4.00	40.00	-0.91	0.09
Reliability		.90						

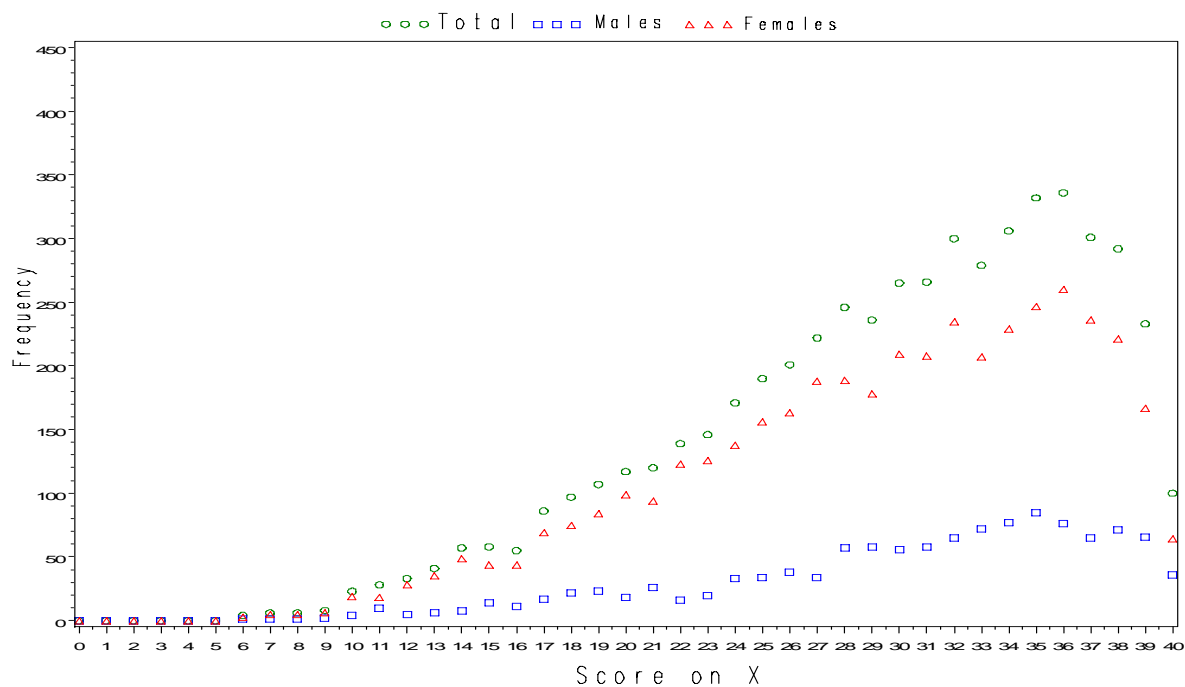


Figure 1. Raw score frequencies on X.

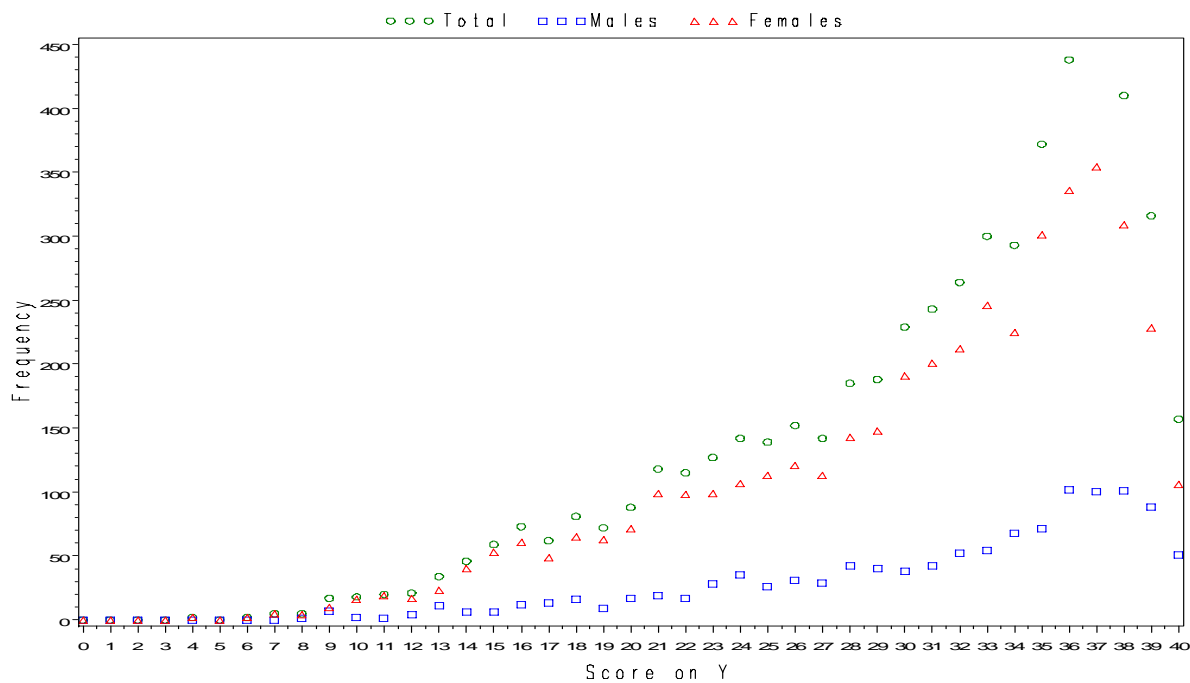


Figure 2. Raw score frequencies on Y.

As illustrated in Figure 3, for scores on Y that are less than 22, the equating function for males was slightly lower than the equating functions for females and for the total group (which is dominated by females). For scores at or above 22, the differences between equating functions were small. As can be seen in Figure 4, the RMSD never exceeded the SDTM. The RMSD was never larger than 0.05, whereas the SDTM was 0.07 ($\text{SDTM} = \text{DTM}/\sigma_X \approx .5/7.17 \approx .07$). The REMSD (not shown) was also small, with a value of 0.01.

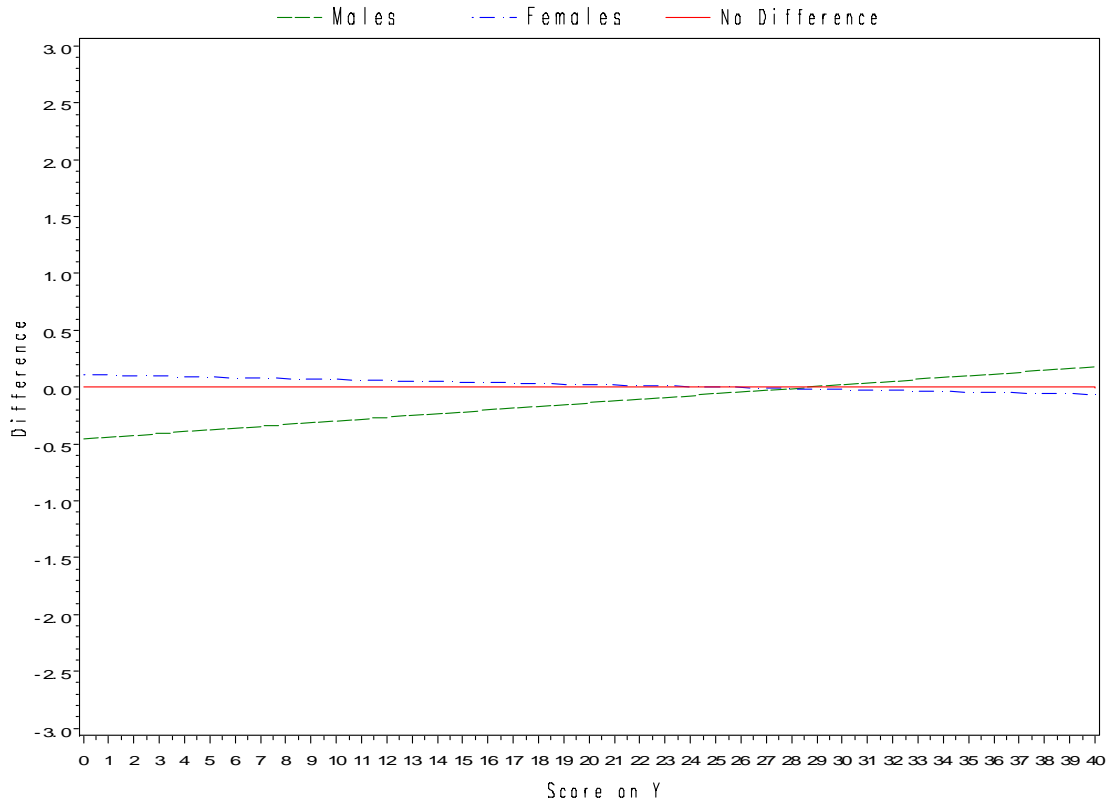


Figure 3. Difference in equated scores between males and total group ($e_{P0} - e_P$) and females and total group ($e_{P1} - e_P$) equating functions.

Based on the estimated values of the RSMD and REMSD, the equating function is likely to be interpreted as population-invariant. The confidence with which this conclusion can be drawn based on the estimated RMSD and REMSD depends to a large degree on how accurate these estimates are. This brings us to the standard errors of the two measures of population

invariance. More specifically, confidence intervals can be constructed based on the standard errors. If the DTM is larger than the upper bound of the confidence interval of the RMSD or REMSD, the equating function can be viewed as population-invariant (at a particular score point for the RMSD). On the other hand, a DTM that is smaller than the upper bound of these confidence intervals casts some doubt on the population invariance of the equating function, even if the actual estimated RMSD or REMSD is smaller than the DTM.

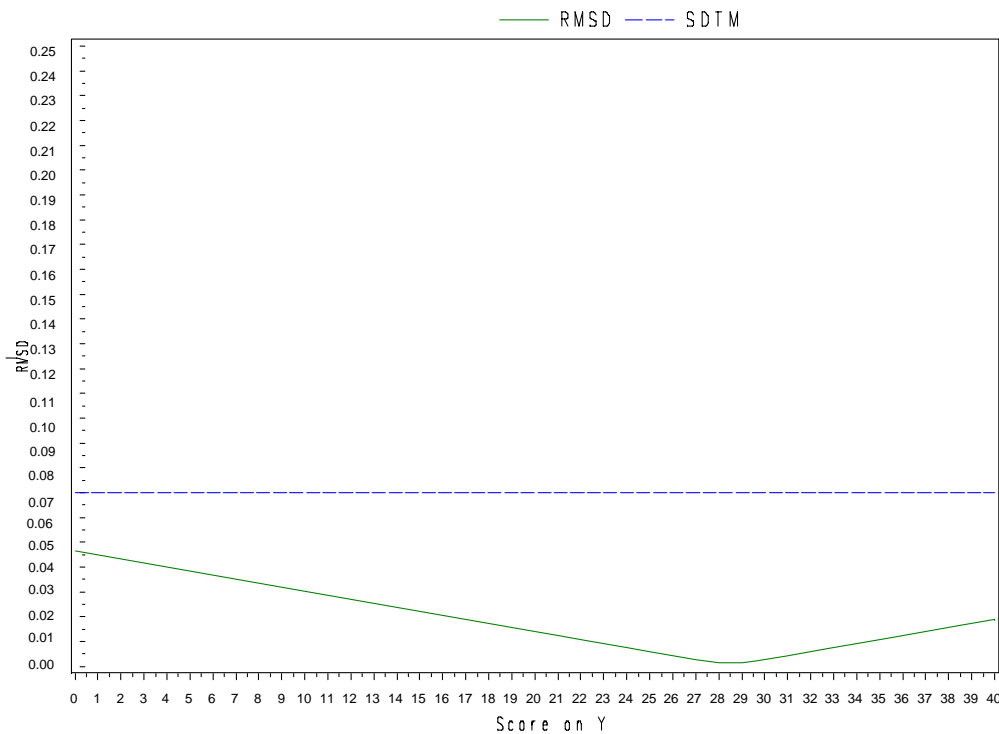


Figure 4. RMSD for the linear equating function.

Standard errors were computed with the two methods presented in this report: the delta method, which provides asymptotic standard errors; and the grouped jackknife method, which is a sampling-based method. For the latter, results in this section are based on $g = 100$ jackknife groups. This choice for g resulted in fairly stable estimates of the standard errors that could be computed in a reasonable time (less than a minute). With respect to the REMSD, the asymptotic standard error is very similar to the sampling-based standard error, with a rounded value of 0.02

for both methods. With an SDTM of 0.07 and an estimated REMSD of 0.01, we can be rather confident that the true REMSD is smaller than the SDTM and therefore that the equating function is population-invariant over the whole range of the scores.

The asymptotic and sampling-based standard errors for the RMSD are displayed in Figure 5. Both methods give very similar standard errors that range between .01 and .10, where the absolute difference between the standard errors is never larger than .01. Note that for both methods, the estimated standard errors show a sudden dip in the score range 27 to 30. This may be because the RMSD becomes very small in that range (see Figure 4). As mentioned before, the RMSD is not differentiable at 0, which is a regularity condition for both methods. Contrary to the results for the REMSD, taking into account the standard errors for the RMSDs casts some doubts on the population invariance of the equating function, because the DTM of 0.07 is smaller than the upper bound of the confidence interval of the RMSD for most score points.

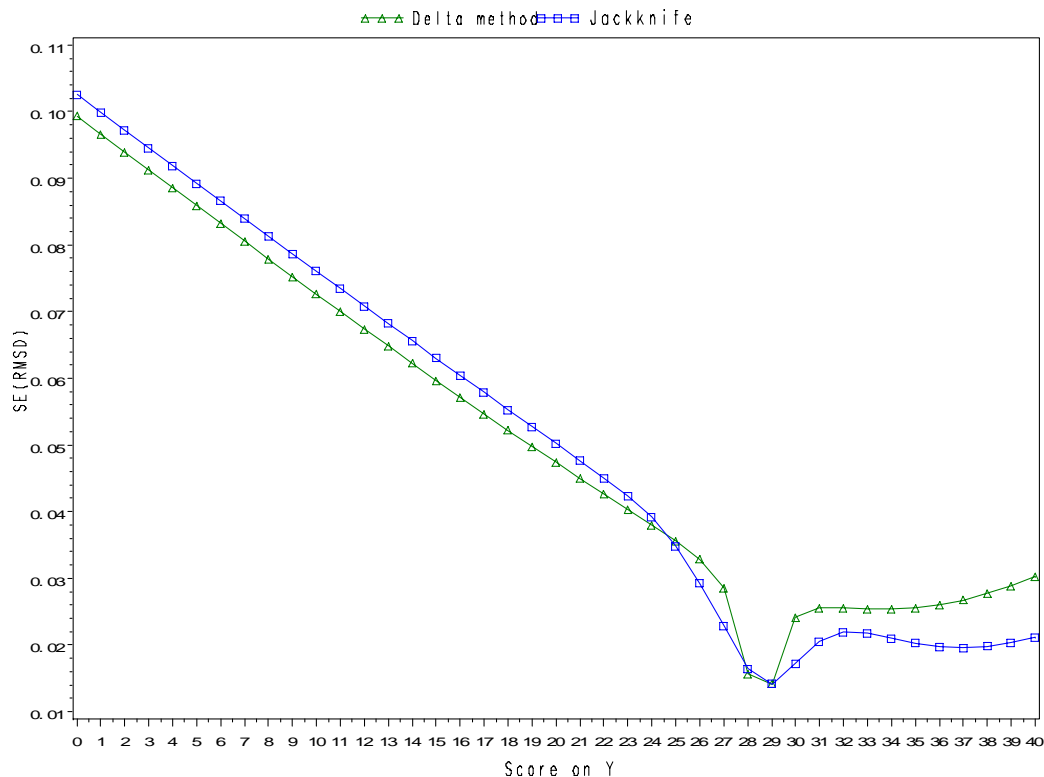


Figure 5. Asymptotic and sampling-based standard errors for the RMSD.

Figure 6 displays the asymptotic and sampling-based standard errors for the RMSD for the different sample sizes. For all sample sizes, both types of standard errors are very close to each other. A similar result was obtained for the REMSD; the asymptotic standard errors amounted to 0.10, 0.06, and 0.04 for sample sizes of 500, 1000, and 2500, respectively, and the sampling-based standard errors amounted to 0.12, 0.06, and 0.05.

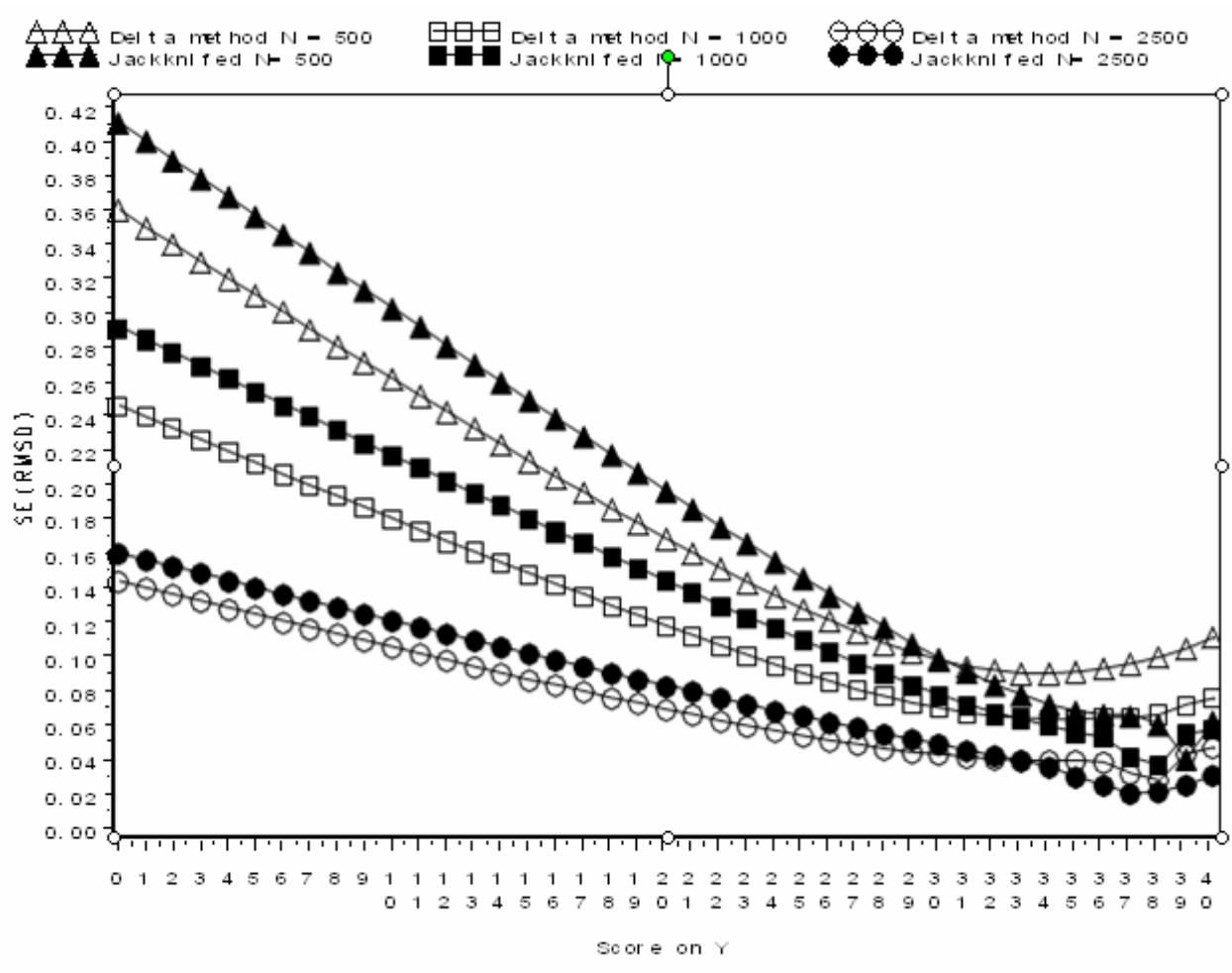


Figure 6. Asymptotic and sampling-based standard errors for the RMSD for different sample sizes.

Discussion

The purpose of this paper was to obtain and examine the standard errors of the RMSD and REMSD, two measures of population invariance of the equating function. Using the delta method, we derived the asymptotic standard errors of the RMSD and REMSD in the context of

linear equating in an EG equating design. In addition, we obtained sampling-based standard errors using a grouped jackknife procedure. Assessing the accuracy of the RMSD and the REMSD in this way is important in determining their reliability in drawing conclusions such as whether or not a particular equating function can be applied to the whole population or only to a subset of the population.

Using a large dataset obtained from an operational administration of a licensure exam, the delta method and the grouped jackknife method both indicated that caution is merited when using RMSD and REMSD to conclude that an equating function is population-invariant. Even though both the RMSD and REMSD were consistently below the SDTM, when 95% confidence intervals were constructed for the RMSD, it exceeded the SDTM at most score points.

The asymptotic standard errors of the delta method were very close to the sampling-based standard errors of the grouped jackknife method. Hence, sampling-based standard errors could be used as a viable alternative to asymptotic standard errors. This is an important result because the derivations for the delta method become quite tedious for designs more complex than the EG design and for equating functions more complex than the linear function. For sampling-based methods, these extensions are straightforward.

References

- Bishop, Y. M. M., Feinberg, S. E., & Holland, P. W. (1975). *Discrete multivariate analysis: Theory and practice*. Cambridge, MA: MIT Press.
- Dorans, N.J. (2003). Population invariance of score linking: *Theory and applications to Advanced Placement Program examinations* (ETS RR-03-27). Princeton, NJ: ETS
- Dorans, N. J., & Feigenbaum, M. D. (1994). Equating issues engendered by changes to the SAT and PSAT/NMSQT. In *Technical issues related to the introduction of the new SAT and PSAT/NMSQT* (ETS Research Memorandum No. RM-94-10). Princeton, NJ: ETS.
- Dorans, N. J., & Holland, P. W. (2000). Population invariance and the equatability of tests: Basic theory and the linear case. *Journal of Educational Measurement*, 37, 281–306.
- Dorans, N. J., Holland, P. W., Thayer, D. T., & Tateneni, K. (2003). Invariance of score linking across gender groups for three Advanced Placement program examinations. In N. J. Dorans (Ed.), *Population invariance of score linking: Theory and applications to Advanced Placement Program examinations* (ETS Research Rep. No. RR-03-27, pp. 79–118). Princeton, NJ: ETS.
- Efron, B. (1982). *The jackknife, the bootstrap and other resampling plans*. Philadelphia: Society for Industrial Mathematics.
- Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. New York: Chapman & Hall.
- Kendall, M., & Stuart, A. (1977). *The advanced theory of statistics* (4th ed., Vol. 1). New York: Macmillan.
- Kolen, M. J. (1985). Standard errors of Tucker equating. *Applied Psychological Measurement*, 9(2), 209-223.
- Kolen, M. J., & Brennan, R. J. (2004). *Test equating: Methods and practices* (2nd ed.) New York: Springer-Verlag.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Erlbaum.
- Moses, T. P. (2006). *Using the kernel method of test equating for estimating the standard errors of population invariance measures* (ETS Research Rep. No. RR-06-20). Princeton, NJ: ETS.
- Rao, C.R. (1973). *Linear statistical inference and applications* (2nd ed.). New York: Wiley.

- von Davier, A. A., Holland, P. W., & Thayer, D. T. (2004). The chain and post-stratification methods for observed-score equating: Their relationship to population invariance. *Journal of Educational Measurement*, 41(1), 15-32.
- von Davier, A.A., & Wilson, C. (2006). Population invariance of IRT true score equating. In A.A. von Davier & M. Liu (Eds.), *Population invariance of test equating and linking: Theory extension and applications across exam* (ETS Research Memorandum No. RM-94-10). Princeton, NJ: ETS.
- Yang, W. L., Dorans, N. J., & Tateneni, K. (2003). Effect of sample selection on Advanced Placement multiple-choice score to composite score linking. In N. J. Dorans (Ed.), *Population invariance of score linking: Theory and applications to Advanced Placement Program examinations* (ETS Research Rep. No. RR-03-27, pp. 57-78). Princeton, NJ: ETS.

Appendix

This appendix provides the entries for $\mathbf{J}_f(\boldsymbol{\pi})$, that is, the first derivatives of the RMSD and REMSD with respect to the parameters in the parameter vector $\boldsymbol{\pi}$ (i.e., μ_{X_0} , $\sigma_{X_0}^2$, μ_{Y_0} , $\sigma_{Y_0}^2$, μ_{X_1} , $\sigma_{X_1}^2$, μ_{Y_1} , and $\sigma_{Y_1}^2$). Specifically, first, two key pieces of the partial derivatives are obtained. These are the partial derivatives of e_P and e_{P_c} with respect to $\boldsymbol{\pi}$. Next, using these key pieces, the partial derivatives of the RMSD are obtained. And finally, the partial derivatives of the REMSD are obtained.

To begin, recall from (1) and (2) that the RMSD and REMSD were defined as

$$\text{RMSD}(y) = \frac{\sqrt{\sum_c w_c [e_{P_c}(y) - e_P(y)]^2}}{\sigma_{XP}}$$

and

$$\text{REMSD} = \frac{\sqrt{\sum_c w_c E_P \left\{ [e_{P_c}(Y) - e_P(Y)]^2 \right\}}}{\sigma_{XP}},$$

respectively. Hence, to obtain the derivatives of the RMSD and REMSD with respect to $\boldsymbol{\pi}$, the derivatives of the linear equating functions for the subgroups (e_{P_c}) and population (e_P) are also needed.

For e_P , the means are obtained from the subpopulation means as

$$\mu_{XP} = w_{X0}^* \mu_{XP_0} + w_{X1}^* \mu_{XP_1}, \quad (\text{A.1})$$

which illustrates that the population mean is a sum of the weighted subgroup means and in which the weights (w_{X0}^* and w_{X1}^*) are defined as the subgroup proportions in the population for test X .

Additionally, σ_{XP}^2 is defined as

$$\sigma_{XP}^2 = w_{X0}^* \sigma_{XP0}^2 + w_{X1}^* \sigma_{XP1}^2 + w_{X0}^* w_{X1}^* (\mu_{XP0} - \mu_{XP1})^2 \quad (\text{A.2})$$

which illustrates that the population variance is the sum of the weighted subgroup variances and the weighted, squared difference between the subgroup means (Kolen, 1985).

Similar relationships hold for test Y, but instead, w_{X0}^* and w_{X1}^* for test X are replaced with w_{Y0}^* and w_{Y1}^* for test Y, σ_X^2 is replaced with σ_Y^2 , and μ_X is replaced with μ_Y .

Using these relationships, the partial derivatives of e_P with respect to π are

$$\frac{\partial e_P}{\partial \mu_{X0}} = \left(\frac{w_{X0}^* w_{X1}^* (\mu_{XP0} - \mu_{XP1})}{\sigma_{XP}} \right) \left(\frac{y - \mu_Y}{\sigma_Y} \right) + w_{X0}^*, \quad (\text{A.2})$$

which can be simplified to

$$\frac{\partial e_P}{\partial \mu_{X0}} = \left(\frac{w_{X0}^* w_{X1}^* (\mu_{XP0} - \mu_{XP1})}{\sigma_{XP}} \right) z_{YP} + w_{X0}^*, \quad (\text{A.3})$$

where

$$z_{YP} = \frac{y - \mu_{YP}}{\sigma_{YP}}, \quad (\text{A.4})$$

$$\frac{\partial e_P}{\partial \mu_{X1}} = - \left(\frac{w_{X0}^* w_{X1}^* (\mu_{XP0} - \mu_{XP1})}{\sigma_{XP}} \right) z_{YP} + w_{X1}^*, \quad (\text{A.5})$$

$$\frac{\partial e_P}{\partial \mu_{Y0}} = - \left(\frac{\sigma_{XP} w_{Y0}^* w_{Y1}^* (\mu_{YP0} - \mu_{YP1})}{\sigma_{YP}^2} \right) z_{YP} - \frac{\sigma_{XP} w_{Y0}^*}{\sigma_{YP}}, \quad (\text{A.6})$$

$$\frac{\partial e_P}{\partial \mu_{Y1}} = \left(\frac{\sigma_{XP} w_{Y0}^* w_{Y1}^* (\mu_{YP0} - \mu_{YP1})}{\sigma_{YP}^2} \right) z_{YP} - \frac{\sigma_{XP} w_{Y1}^*}{\sigma_{YP}}, \quad (\text{A.7})$$

$$\frac{\partial e_P}{\partial \sigma_{XP0}^2} = \left(\frac{w_{X0}^*}{2\sigma_{XP}} \right) z_{YP}, \quad (\text{A.8})$$

$$\frac{\partial e_P}{\partial \sigma_{XP1}^2} = \left(\frac{w_{X1}^*}{2\sigma_{XP}} \right) z_{YP}, \quad (\text{A.9})$$

$$\frac{\partial e_P}{\partial \sigma_{YP0}^2} = - \left(\frac{w_{Y0}^* \sigma_{XP}}{2\sigma_{YP}^2} \right) z_{YP}, \quad (\text{A.10})$$

and

$$\frac{\partial e_P}{\partial \sigma_{YP1}^2} = - \left(\frac{w_{Y1}^* \sigma_{XP}}{2\sigma_{YP}^2} \right) z_{YP}. \quad (\text{A.11})$$

For e_{P_c} , the partial derivatives with respect to π are

$$\frac{\partial e_{P_c}}{\partial \mu_{XP_c}} = 1, \quad (\text{A.12})$$

$$\frac{\partial e_{P_c}}{\partial \mu_{YP_c}} = - \frac{\sigma_{XP_c}}{\sigma_{YP_c}}, \quad (\text{A.13})$$

$$\frac{\partial e_{P_c}}{\partial \sigma_{XP_c}^2} = \left(\frac{1}{2\sigma_{XP_c}} \right) z_{YP_c}, \quad (\text{A.14})$$

$$\frac{\partial e_{P_c}}{\partial \sigma_{YP_c}^2} = - \left(\frac{\sigma_{XP_c}}{2\sigma_{YP_c}^2} \right) z_{YP_c}, \quad (\text{A.15})$$

$$\frac{\partial e_{P0}}{\partial \mu_{XP1}} = \frac{\partial e_{P0}}{\partial \mu_{YP1}} = \frac{\partial e_{P0}}{\partial \sigma_{XP1}^2} = \frac{\partial e_{P0}}{\partial \sigma_{YP1}^2} = 0, \quad (\text{A.16})$$

and

$$\frac{\partial e_{P1}}{\partial \mu_{XP0}} = \frac{\partial e_{P1}}{\partial \mu_{YP0}} = \frac{\partial e_{P1}}{\partial \sigma_{XP0}^2} = \frac{\partial e_{P1}}{\partial \sigma_{YP0}^2} = 0. \quad (\text{A.17})$$

Partial Derivatives of the RMSD With Respect to π

Using the previously mentioned partial derivatives and the quotient and chain rules of derivations, the partial derivative of the RMSD with respect to μ_{X0} is obtained

$$\begin{aligned}
\frac{\partial \text{RMSD}}{\partial \mu_{XP_0}} = & - \left(\frac{w_{X0}^* w_{X1}^* (\mu_{XP_0} - \mu_{XP_1})}{\sigma_{XP}^2} \right) \text{RMSD} + \left(\frac{1}{\sigma_{XP} \text{RMSD}} \right) \times \\
& \left[\sum_c \left[\frac{w_c [e_{P_c}(y) - e_P(y)]}{\sigma_{XP}} \right] \times \left[I_0^{(c)} - \left(\frac{w_{X0}^* w_{X1}^* (\mu_{XP_0} - \mu_{XP_1})}{\sigma_{XP}} \right) z_{YP} - w_{X0}^* \right] \right],
\end{aligned} \tag{A.18}$$

where $I_{c'}^{(c)}$ is a dummy variable, in which the subscript c' is a dummy coded symbol for the subgroup for which the function is being derived and the superscript c is a dummy coded symbol for the subgroup considered in the summation. That is, define

$$I_{c'}^{(c)} = \begin{cases} 0 & \text{if } c \neq c' \\ 1 & \text{if } c = c' \end{cases} \tag{A.19}$$

To simplify the expression, let

$$\Phi \equiv \frac{1}{\text{RMSD}}. \tag{A.20}$$

Another way to conceptualize Φ is as the similarity between the subpopulation linking functions and the overall linking function across the score scale.

Additionally, define the weighted, standardized difference between the equating functions for subpopulation c and the total population as (Δ_c)

$$\Delta_c \equiv \frac{w_c [e_{P_c}(y) - e_P(y)]}{\sigma_{XP}}. \tag{A.21}$$

Substituting (A.20)-(A.21) into (A.18) yields the following

$$\begin{aligned} \frac{\partial \text{RMSD}}{\partial \mu_{XP_0}} = & - \left(\frac{w_{X0}^* w_{X1}^* (\mu_{XP_0} - \mu_{XP_1})}{\sigma_{XP}^2} \right) \text{RMSD} + \left(\frac{\Phi}{\sigma_{XP}} \right) \times \\ & \left[\sum_c \Delta_c \times \left[I_0^{(c)} - \left(\frac{w_{X0}^* w_{X1}^* (\mu_{XP_0} - \mu_{XP_1})}{\sigma_{XP}} \right) z_{YP} - w_{X0}^* \right] \right], \end{aligned} \quad (\text{A.22})$$

which illustrates that the derivative of the RMSD is primarily a function of the RMSD itself; relative sample proportion ; the precision of the equating functions (Φ); the weighted, standardized difference between the equating functions (Δ_c); and the standardized scores on the form that is being equated (z_Y).

Using a similar approach, the following partial derivatives are also obtained:

$$\begin{aligned} \frac{\partial \text{RMSD}}{\partial \mu_{XP_1}} = & \left(\frac{w_{X0}^* w_{X1}^* (\mu_{XP_0} - \mu_{XP_1})}{\sigma_{XP}^2} \right) \text{RMSD} + \left(\frac{\Phi}{\sigma_{XP}} \right) \times \\ & \left[\sum_c \Delta_c \times \left[I_1^{(c)} + \left(\frac{w_{X0}^* w_{X1}^* (\mu_{XP_0} - \mu_{XP_1})}{\sigma_{XP}} \right) z_{YP} - w_{X1}^* \right] \right], \end{aligned} \quad (\text{A.23})$$

$$\begin{aligned} \frac{\partial \text{RMSD}}{\partial \mu_{YP_0}} = & \left(\frac{\Phi}{\sigma_{XP}} \right) \times \\ & \left[\sum_c \Delta_c \times \left[I_0^{(c)} \frac{-\sigma_{XP_0}}{\sigma_{YP_0}} + \left(\frac{\sigma_{XP} w_{Y0}^* w_{Y1}^* (\mu_{YP_0} - \mu_{YP_1})}{\sigma_{YP}^2} \right) z_{YP} + \frac{\sigma_{XP} w_{Y0}^*}{\sigma_{YP}} \right] \right], \end{aligned} \quad (\text{A.24})$$

$$\begin{aligned} \frac{\partial \text{RMSD}}{\partial \mu_{YP_1}} = & \left(\frac{\Phi}{\sigma_{XP}} \right) \times \\ & \left[\sum_c \Delta_c \times \left[I_1^{(c)} \frac{-\sigma_{XP_1}}{\sigma_{YP_1}} - \left(\frac{\sigma_{XP} w_{Y0}^* w_{Y1}^* (\mu_{YP_0} - \mu_{YP_1})}{\sigma_{YP}^2} \right) z_{YP} + \frac{\sigma_{XP} w_{Y1}^*}{\sigma_{YP}} \right] \right], \end{aligned} \quad (\text{A.25})$$

$$\frac{\partial \text{RMSD}}{\partial \sigma_{XP_0}^2} = - \left(\frac{w_{X0}^*}{2\sigma_{XP}^2} \right) \text{RMSD} + \left(\frac{\Phi}{\sigma_{XP}} \right) \times \left[\sum_c \Delta_c \times \left[\frac{I_0^{(c)} z_{YP_0}}{2\sigma_{XP_0}} - \frac{w_{X0}^* z_{YP}}{2\sigma_{XP}} \right] \right], \quad (\text{A.26})$$

$$\frac{\partial \text{RMSD}}{\partial \sigma_{XP1}^2} = -\left(\frac{w_{X1}^*}{2\sigma_{XP}^2}\right) \text{RMSD} + \left(\frac{\Phi}{\sigma_{XP}}\right) \times \left[\sum_c \Delta_c \times \left[\frac{I_1^{(c)} z_{YP1}}{2\sigma_{XP1}} - \frac{w_{X1}^* z_{YP}}{2\sigma_{XP}} \right] \right], \quad (\text{A.27})$$

$$\frac{\partial \text{RMSD}}{\partial \sigma_{YP0}^2} = \left(\frac{\Phi}{\sigma_{XP}}\right) \times \left[\sum_c \Delta_c \times \left[\left(\frac{w_{Y0}^* \sigma_{XP}}{2\sigma_{YP}^2}\right) z_{YP} - I_0^{(c)} \left(\frac{\sigma_{XP0}}{2\sigma_{YP0}^2}\right) z_{YP0} \right] \right], \quad (\text{A.28})$$

and

$$\frac{\partial \text{RMSD}}{\partial \sigma_{YP1}^2} = \left(\frac{\Phi}{\sigma_{XP}}\right) \times \left[\sum_c \Delta_c \times \left[\left(\frac{w_{Y1}^* \sigma_{XP}}{2\sigma_{YP}^2}\right) z_{YP} - I_1^{(c)} \left(\frac{\sigma_{XP1}}{2\sigma_{YP1}^2}\right) z_{YP1} \right] \right]. \quad (\text{A.29})$$

Partial Derivatives of the REMSD With Respect to π

Similarly, and again using the previously mentioned partial derivatives, the quotient and chain rules of derivations, and the rules of expectations, the partial derivative of the REMSD with respect to μ_{X0} is

$$\begin{aligned} \frac{\partial \text{REMSD}}{\partial \mu_{XP0}} = & -\left(\frac{w_{X0}^* w_{X1}^* (\mu_{XP0} - \mu_{XP1})}{\sigma_{XP}^2}\right) \text{REMSD} + \left(\frac{1}{\sigma_{XP} \text{REMSD}}\right) \times \\ & \left[\sum_c E_P \left\{ \Delta_c \times \left[I_0^{(c)} - \left(\frac{w_{X0}^* w_{X1}^* (\mu_{XP0} - \mu_{XP1})}{\sigma_{XP}}\right) z_{YP} - w_{X0}^* \right] \right\} \right]. \end{aligned} \quad (\text{A.30})$$

To simplify the expression, define Ψ as the inverse of the REMSD

$$\Psi \equiv \frac{1}{\text{REMSD}}, \quad (\text{A.31})$$

which can be conceptualized as the expectation of the similarity between the subpopulation linking functions and the overall linking function across the score scale.

Substituting (A.31) into (A.30) yields

$$\begin{aligned} \frac{\partial \text{REMSD}}{\partial \mu_{XP0}} = & - \left(\frac{w_{X0}^* w_{X1}^* (\mu_{XP0} - \mu_{XP1})}{\sigma_{XP}^2} \right) \text{REMSD} + \left(\frac{\Psi}{\sigma_{XP}} \right) \times \\ & \left[\sum_c E_P \left\{ \Delta_c \times \left[I_0^{(c)} - \left(\frac{w_{X0}^* w_{X1}^* (\mu_{XP0} - \mu_{XP1})}{\sigma_{XP}} \right) z_{YP} - w_{X0}^* \right] \right\} \right]. \end{aligned} \quad (\text{A.32})$$

Using a similar approach, the following partial derivatives of the REMSD are also obtained:

$$\begin{aligned} \frac{\partial \text{REMSD}}{\partial \mu_{XP1}} = & \left(\frac{w_{X0}^* w_{X1}^* (\mu_{XP0} - \mu_{XP1})}{\sigma_{XP}^2} \right) \text{REMSD} + \left(\frac{\Psi}{\sigma_{XP}} \right) \times \\ & \left[\sum_c E_P \left\{ \Delta_c \times \left[I_1^{(c)} + \left(\frac{w_{X0}^* w_{X1}^* (\mu_{XP0} - \mu_{XP1})}{\sigma_{XP}} \right) z_{YP} - w_{X1}^* \right] \right\} \right], \end{aligned} \quad (\text{A.33})$$

$$\begin{aligned} \frac{\partial \text{REMSD}}{\partial \mu_{YP0}} = & \left(\frac{\Psi}{\sigma_{XP}} \right) \times \\ & \left[\sum_c E_P \left\{ \Delta_c \times \left[I_0^{(c)} \frac{-\sigma_{XP0}}{\sigma_{YP0}} + \left(\frac{\sigma_{XP} w_{Y0}^* w_{Y1}^* (\mu_{YP0} - \mu_{YP1})}{\sigma_{YP}^2} \right) z_{YP} + \frac{\sigma_{XP} w_{Y0}^*}{\sigma_{YP}} \right] \right\} \right], \end{aligned} \quad (\text{A.34})$$

$$\begin{aligned} \frac{\partial \text{REMSD}}{\partial \mu_{YP1}} = & \left(\frac{\Psi}{\sigma_{XP}} \right) \times \\ & \left[\sum_c E_P \left\{ \Delta_c \times \left[I_1^{(c)} \frac{-\sigma_{XP1}}{\sigma_{YP1}} - \left(\frac{\sigma_{XP} w_{Y0}^* w_{Y1}^* (\mu_{YP0} - \mu_{YP1})}{\sigma_{YP}^2} \right) z_{YP} + \frac{\sigma_{XP} w_{Y1}^*}{\sigma_{YP}} \right] \right\} \right], \end{aligned} \quad (\text{A.35})$$

$$\begin{aligned} \frac{\partial \text{REMSD}}{\partial \sigma_{XP0}^2} = & - \left(\frac{w_{X0}^*}{2\sigma_{XP}^2} \right) \text{REMSD} + \left(\frac{\Psi}{\sigma_{XP}} \right) \times \left[\sum_c E_P \left\{ \Delta_c \times \left[\frac{I_0^{(c)} z_{YP0}}{2\sigma_{XP0}} - \frac{w_{X0}^* z_{YP}}{2\sigma_{XP}} \right] \right\} \right], \end{aligned} \quad (\text{A.36})$$

$$\frac{\partial \text{REMSD}}{\partial \sigma_{XP1}^2} = -\left(\frac{w_{X1}^*}{2\sigma_{XP}^2}\right) \text{REMSD} + \left(\frac{\Psi}{\sigma_{XP}}\right) \times \left[\sum_c E_P \left\{ \Delta_c \times \left[\frac{I_1^{(c)} z_{YP1}}{2\sigma_{XP1}} - \frac{w_{X1}^* z_{YP}}{2\sigma_{XP}} \right] \right\} \right], \quad (\text{A.37})$$

$$\frac{\partial \text{REMSD}}{\partial \sigma_{YP0}^2} = \left(\frac{\Psi}{\sigma_{XP}}\right) \times \left[\sum_c E_P \left\{ \Delta_c \times \left[\left(\frac{w_{Y0}^* \sigma_{XP}}{2\sigma_{YP}^2} \right) z_{YP} - I_0^{(c)} \left(\frac{\sigma_{XP0}}{2\sigma_{YP0}^2} \right) z_{YP0} \right] \right\} \right], \quad (\text{A.38})$$

and

$$\frac{\partial \text{REMSD}}{\partial \sigma_{YP1}^2} = \left(\frac{\Psi}{\sigma_{XP}}\right) \times \left[\sum_c E_P \left\{ \Delta_c \times \left[\left(\frac{w_{Y1}^* \sigma_{XP}}{2\sigma_{YP}^2} \right) z_{YP} - I_1^{(c)} \left(\frac{\sigma_{XP1}}{2\sigma_{YP1}^2} \right) z_{YP1} \right] \right\} \right]. \quad (\text{A.39})$$